AD_____

Award Number: W81XWH-06-1-0220

TITLE: Identification of Ovarian Cancer Susceptibility Genes Involved in Stromal-Epithlial Interactions

PRINCIPAL INVESTIGATOR: Georgia Chenevix-Trench

CONTRACTING ORGANIZATION: Queensland Institute of Medical Research
                           Herston QLD 4029 Australia

REPORT DATE: February 2008

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
               Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
                        Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01-02-2008	Annual	1 Feb 2007 – 31 Jan 2008

4. TITLE AND SUBTITLE

Identification of Ovarian Cancer Susceptibility Genes Involved in Stromal-Epithlial Interactions

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-06-1-0220

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Georgia Chenevix-Trench with Izhak Haviv, Penny Webb, Dorota Gertig, David Duffy, Paul Pharoah, Alice Whittemore, Ellen Goode, Joellen Schildkraut, Hoda Anton-Culver and Leigh Pearce

Email: georgiaT@qimr.edu.au

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Queensland Institute of Medical Research
Herston QLD 4029 Australia

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The purpose of this project is to identify ovarian cancer susceptibility genes involved in stromal- epithelial cross talk. We have now genotyped 2138 samples for 1536 tagging, non-synonymous and miRNA binding site SNPs in 174 genes. Following Quality Control exclusions, the final dataset comprised 1839 samples (675 cases and 1164 controls) with genotype information for 1292 SNPs in 174 genes. We are using P(trend) values to select 25-30 SNPs for independent validation in ~2000 cases and ~4000 controls from other sites. Four of these SNPs will also be genotyped by the whole of the Ovarian Cancer Association Consortium (OCAC) (~4500 case and ~6500 controls). Currently our most significant finding is for a SNP in the podocalyxin-like (PODXL) gene for which the P(trend) = 0.0001, with a Positive Predictive Value of 33 %, and a Homozygote OR = 1.75 (95% CI 1.28-3.38).

15. SUBJECT TERMS
risk, ovarian cancer, polymorphisms, genes, stromal-epithelial interactions

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	16	**19b. TELEPHONE NUMBER** (include area code)
U	U	U			

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

INTRODUCTION

We propose that subtle variation in the expression or function of genes expressed as a consequence of interactions between ovarian cancer cells and the host micro-environment could contribute to susceptibility to ovarian cancer. This idea is novel because this class of genes has not previously been tested for a role in ovarian cancer susceptibility. Our approach, and our choice of candidate genes, is based on extensive preliminary data we have accumulated from co-culture of fibroblast and epithelial ovarian cells. Our original aim was to identify all non-synonymous coding and putative promoter SNPs in 60 candidate genes highlighted by our analysis of cross talk between fibroblast and epithelial elements of ovarian tumors, as well as a set of haplotype tagging SNPs in 20 of these co-culture regulated genes which are altered in expression in serous tumours, compared with normal ovarian surface epithelial cells. However, since the start of this project we have acquired an Illumina Bead Station and so we genotyped 1536 SNPs in the first stage, allowing us to genotype potentially functional as well as tagging SNPs in 174 genes of interest in 773 cases with invasive, serous ovarian adenocarcinoma, and 1365 controls. This task has been completed and will be followed by independent validation of the most significant associations using a replication set of at least 2,100 cases with serous ovarian adenocarcinoma and 3,600 controls. Finally, we will look for the putative functional SNPs in these genes, and evaluate their function *in vitro*.

BODY

The statement of work was altered in December 2006 because we changed genotyping platforms in order to genotype many more SNPs, but with an altered the time frame. The tasks below are from the new SOW.

Task 1. In silico **identification of SNPs in candidate genes (months 1-9)**

1. *identification of 174 candidate genes involved in cross talk*

The original application proposed genotyping of candidate genes based on a series of *in vitro* experiments involving co-culture of ovarian epithelial and theca fibroblast cells. The genes were further prioritized based on elevated expression in two published ovarian cancer expression profiling studies, as well as an in house expression profile and we then generated a list of 255 candidate genes of interest.

2. *identification of 1536 tagging SNPs, nsSNPs and SNPs in putative microRNA binding sites in these 174 genes*

With Drs Ellen Goode and David Rider at the Mayo Clinic, and Illumina Inc., we then generated a list of SNPs within 5 kb of these 255 genes (58,114 SNPs in total). We then used the binning algorithm of LDSelect to identify 4567 tagSNPs among these, with (r2) $\geq$ 0.8 and minor allele frequencies (MAFs) > 0.05, using data from a variety of sources. Then we prioritized the list to 166 genes based on known function and the number of bins in each gene (excluding genes with a large number of bins), in an attempt to reduce the list to ~1500 SNPs.

We then requested from Illumina Inc the design scores for all SNPs within 5kb of these 166 genes and picked the best tagSNP in each bin (or two tagSNPs if there are >10 tagging SNPs in a bin and none had an optimal design score).We also used www.patrocles.org to identify SNPs (with MAFs $\geq$ 0.05) in microRNA binding sites within these genes, and added nsSNPs (with MAFs $\geq$ 0.05) from the public

databases to the potential SNP list. This identified 170 miRNA binding sites and nsSNPs with Illumina design scores > 0.6 in these 166 genes. In total this gave 1410 tagSNPs, miRNA binding site SNPs and nsSNPs, and so the list was supplemented by tag and supplemental SNPs in another 12 candidate genes, bringing the number of genes represented in the final list to 174, in which there were 1509 SNPs meeting the above criteria (some of the original 174 candidate genes had no appropriate SNPs in them). In order to reach the final total of 1536 SNPs for the Illumina OPA, the MAF of the supplemental SNPs was dropped to 0.01. The final list of 1536 SNPs included 106 supplemental SNPs and 1430 tagSNPs. The Illumina OPA for these 1536 SNPs was ordered in December 2006, and received early in February 2007.

Task 2.   Genotyping of 900 cases and 1200 controls for 1536 SNPs using the Illumina Goldengate Assay (months 10-15)

While the design of the Illumina OPA was underway we completed the extraction and Quality Control of 1350 case and 1100 controls DNAs from the Australian Ovarian Cancer Study (AOCS), and the making of plates for Goldengate genotyping using cases and controls from both the AOCS and the Australian Cancer Study.

We have now genotyped 2138 samples for 1536 SNPs in 174 genes. There were 773 invasive serous cases from the Australian Ovarian Cancer Study (527), Australian Cancer Study (121) and Mayo Clinic (125), with 1365 controls from the same sources (893, 411 and 61 respectively). Insufficient DNA was available from the AOCS to achieve our original aim of genotyping 900 invasive serous cases, but additional power was obtained by using a larger number of controls.

Plates were prepared containing randomly mixed cases and controls, with two duplicated samples and one blank per plate.   The Golden Gate assay was performed according to the manufacturer's instructions. Following completion of the assay for all 23 plates, analysis was carried out using Illumina BeadStudio software version 3.1.0.0. The following quality control measures were implemented:

The original raw dataset contained genotype information for 2208 samples and 1536 SNPs. Following automatic clustering, SNPs were ranked using their "GenTrain" score, number between 0 and 1 indicating how well the samples clustered for this locus. SNPs with a low score were checked manually and re-clustered if possible.   Subsequently all SNPs were checked for clustering quality.

Next, SNPs were filtered based on call rate with a call rate > 95% deemed as acceptable. Additional filter steps included removal of SNPs with a minor allele frequency of zero.   Hardy Weinberg equilibrium was also tested for each SNP, and only those that passed a low threshold, with a p value > 0.0001, were included. SNPs with two or more discrepancies between duplicate pairs were excluded.

For sample quality control, a call rate threshold of 95% was used so that samples that failed for $\geq$95% SNPs were excluded which reduced the number of samples from 2208 to 2145.   Analysis of signal intensities across all plates revealed three plates (#16-18) with low intensity, just prior to the annual service of the laser. A separate analysis looking at call rates and concordance for each plate showed that these same plates failed quality control thresholds (Figure 1) and so they were omitted from further analysis.

Figure: Call frequencies for SNPs in each plate genotyped
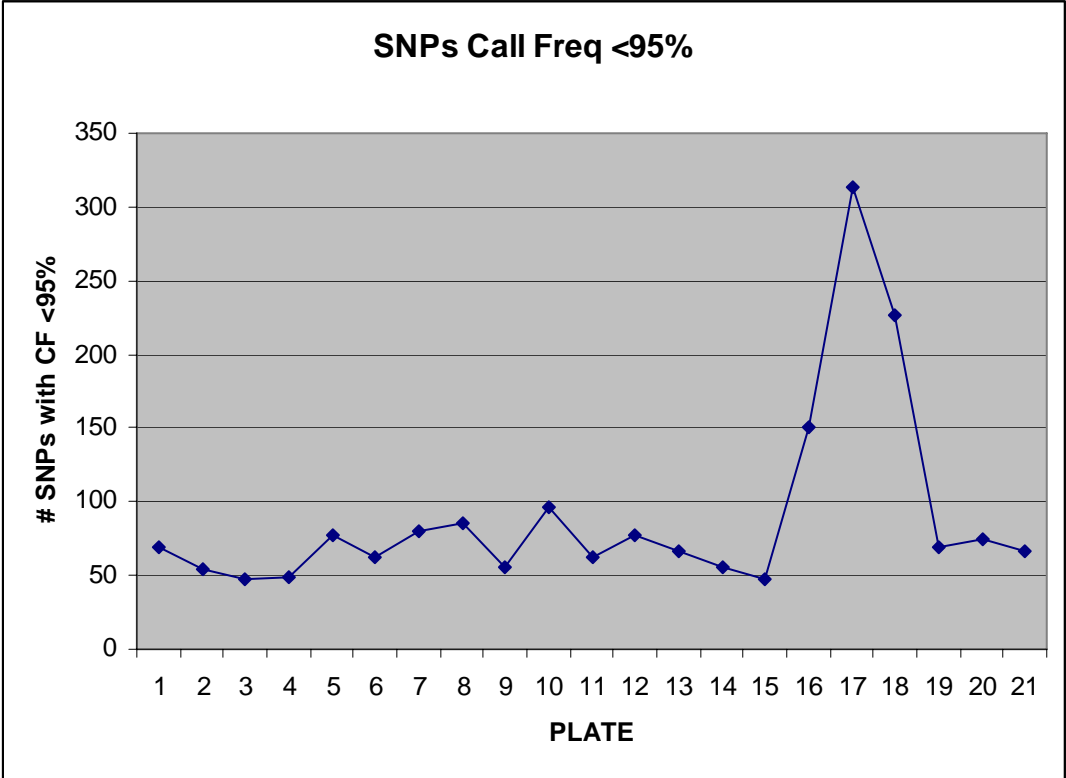
The final dataset therefore comprised 1839 samples (675 cases and 1164 controls) with genotype information for 1292 SNPs in 174 genes (Table 1). An analysis using the PLINK software package was then performed.

Table 1: SNP selection and Quality Control

Gene Name	Total SNPs	NOT ON OPA	FUNCTIONAL SNP	tgSNP	PASSED QC	FAILED QC
ADAM8	6	0	1	5	5	1
ANKRD1	9	0	0	9	9	0
AREG	6	4	0	2	2	0
BCL2L1	4	0	0	4	4	0
BMP1	20	6	0	14	10	4
BMP4	4	0	0	4	2	2
BPNT1	7	2	0	5	5	0
BST2	2	0	1	1	0	2
BUB1	1	0	0	1	1	0
C3	67	34	5	28	26	7
CCL11	8	2	0	6	5	1
CCL13	4	0	0	4	4	0
CCL7	11	3	0	8	7	1
CCND2	25	1	0	24	20	4
CD24	2	0	1	1	1	1
CD44	58	4	1	53	48	6
CFLAR	3	0	0	3	3	0
cig5	7	0	0	7	7	0

CRLF3	2	0	1	1	2	0
CSF1	12	0	1	11	10	2
CTGF	5	0	0	5	4	1
CTSK	3	1	0	2	2	0
CXCL1	4	0	0	4	4	0
CXCL14	10	1	0	9	9	0
CXCL3	1	0	0	1	1	0
CXCL6	1	0	0	1	1	0
CXCL9	8	0	0	8	8	0
CXCR6	5	2	0	3	2	1
CYC1	2	0	0	2	2	0
CYR61	8	0	0	8	7	1
DAB2	10	1	2	7	6	3
DCN	7	2	1	4	3	2
DDR2	29	2	0	27	21	6
DKK1	2	0	0	2	2	0
DLG7	4	1	0	3	3	0
DPP4	12	0	0	12	11	1
DSC3	32	5	2	25	24	3
DUSP5	2	0	0	2	2	0
EGF	18	3	3	12	12	3
EGR2	3	0	0	3	3	0
EIF4EBP2	4	0	1	3	3	1
ERBB3	5	0	0	5	2	3
FGF2	20	0	0	20	18	2
FLT3LG	4	0	0	4	4	0
FN1	24	3	1	20	18	3
FOS	4	0	0	4	4	0
FST	6	2	0	4	3	1
G1P2	2	0	1	1	2	0
G1P3	2	0	0	2	2	0
GABARAPL1	10	2	1	7	6	2
GAS1	1	0	0	1	1	0
GATA6	5	0	0	5	5	0
GJB1	4	0	0	4	4	0
GJB2	3	0	1	2	3	0
GPX4	3	2	0	1	1	0
H1F0	4	0	1	3	3	1
HIF1A	11	1	2	8	10	0
HOXB2	3	0	0	3	2	1
ID2	2	0	0	2	2	0
IFI16	12	2	2	8	9	1
IFI35	5	0	0	5	4	1
IFIT1	5	0	0	5	5	0
IFITM1	2	1	0	1	1	0
IFITM2	5	5	0	0	0	0
IGFBP3	18	9	1	8	7	2
IGFBP4	6	0	0	6	5	1
IGFBP5	12	1	1	10	9	2
IL1B	10	2	0	8	8	0

IL1R1	24	6	0	18	15	3
IL6	10	3	1	6	6	1
IL6ST	7	0	2	5	6	1
IL8	1	0	0	1	1	0
INHBA	4	0	0	4	3	1
IRF7	6	2	1	3	3	1
ITGA6	**25**	**3**	**1**	**21**	**18**	**4**
ITGAV	18	0	0	18	17	1
ITGB1	19	3	0	16	14	2
JUN	6	1	1	4	5	0
LAMC1	11	1	2	8	9	1
LCN2	5	1	0	4	3	1
MAPK1	9	0	1	8	7	2
MCM2	4	0	1	3	3	1
MCM6	7	0	0	7	6	1
MEST	5	0	0	5	4	1
MFAP4	2	0	1	1	2	0
MKI67	34	1	14	19	25	8
MMP1	18	2	1	15	13	3
MMP14	13	3	0	10	10	0
MMP2	11	2	0	9	9	0
MMP26	7	0	0	7	6	1
MMP3	**8**	**1**	**1**	**6**	**7**	**0**
MMP7	13	1	1	11	9	3
MMP9	7	2	0	5	4	1
MPI	2	0	0	2	2	0
MX1	27	3	1	23	16	8
NFKB2	3	0	0	3	3	0
NFKBIA	14	4	1	9	8	2
NOTCH3	10	1	3	6	7	2
NT5E	9	0	3	6	8	1
OAS1	5	1	1	3	3	1
OAS3	12	1	2	9	8	3
OGT	3	0	0	3	2	1
OSMR	25	2	2	21	20	3
P4HA2	11	1	0	10	8	2
PANX1	15	0	3	12	14	1
PDGFB	11	1	0	10	4	6
PDGFRB	31	5	1	25	21	5
PLAT	13	3	1	9	7	3
PLAU	5	0	1	4	4	1
PLAUR	32	16	3	13	15	1
PLOD	9	0	1	8	7	2
PLOD2	8	0	0	8	7	1
PODXL	**27**	**2**	**2**	**23**	**20**	**5**
PRKR	10	1	0	9	7	2
PRKRA	4	0	0	4	4	0
PRKRIR	4	0	0	4	4	0
PTEN	11	1	0	10	10	0
PTGES	8	0	0	8	6	2

PTGS1	11	1	1	9	8	2
PTGS2	12	2	1	9	9	1
PTP4A1	2	0	0	2	2	0
PTPN1	9	1	0	8	7	1
PTTG1	7	1	0	6	4	2
RGS2	3	1	0	2	2	0
S100A7	5	0	0	5	3	2
SAT	4	0	0	4	4	0
SELENBP1	6	1	0	5	5	0
SERPINB2	5	0	2	3	5	0
SERPINB7	14	1	1	12	11	2
SERPINE1	10	0	1	9	8	2
SERPING1	8	0	1	7	5	3
SIAT9	10	2	1	7	8	0
SNAI1	6	1	0	5	4	1
SORD	2	0	0	2	1	1
SORT1	8	0	0	8	8	0
SOX9	3	0	0	3	2	1
SPARC	13	2	0	11	11	0
SPP1	6	1	0	5	5	0
SPRY1	4	0	0	4	3	1
SSA1	11	0	0	11	8	3
STAT1	18	0	0	18	15	3
STAT3	9	2	0	7	5	2
STEAP	7	0	0	7	7	0
T1A-2	24	1	0	23	19	4
TACSTD1	10	1	1	8	8	1
TERT	27	10	0	17	11	6
TGFB2	25	3	0	22	18	4
TGFB3	7	0	0	7	6	1
THBS4	18	4	0	14	13	1
TIEG	8	0	0	8	8	0
TIMP1	2	0	0	2	1	1
TIMP3	31	3	1	27	26	2
TNF	7	1	2	4	5	1
TNFAIP2	13	6	0	7	6	1
TNFAIP3	7	2	1	4	5	0
TNFAIP6	6	0	1	5	5	1
TNFRSF12A	5	1	1	3	3	1
TNFRSF1B	35	11	2	22	21	3
TNFSF10	14	2	0	12	10	2
TNFSF7	17	8	0	9	7	2
TNFSF9	3	0	0	3	3	0
TWIST1	1	0	0	1	1	0
TYK2	10	2	0	8	6	2
TYROBP	1	0	0	1	1	0
VDR	25	1	3	21	19	5
VEGF	20	6	0	14	11	3
VEGFC	12	0	0	12	10	2
VIL2	13	1	1	11	9	3

WISP1	30	1	0	29	24	5
WNT10B	4	0	1	3	4	0
WNT2	16	1	0	15	13	2
WNT5A	9	0	0	9	7	2
ZNF354A	4	1	0	3	3	0
TOTALS	1796	260	106	1430	1292	244

Task 3. **Genotyping of the AOCS/ACS test set for additional SNPs by Mass Array and statistical analysis of test set (months 16-21)**

1. *genotyping 900 cases and 1200 controls by Mass Array for 70 SNPs that were not amenable to Illumina genotyping in 13 key genes using 30-plexes*

AOCS and ACS case (including non-serous invasive cases and LMP cases) and control DNAs have been randomly plated in 8 x 384 well plates ready for iPLEX genotyping. We originally selected 174 genes for Golden Gate analysis. Many of these genes contain SNPs of interest that were either not amenable to the Golden Gate assay, or were genotyped on the OPA but failed quality control criteria. The genes of most *a priori* biological interest to us are *CXCL9, CTGF, LCN2, DCN,* and *VIL2,* in which there are 11 SNPs that either could not be designed for the OPA, or failed QC on the OPA. In addition, we will genotype 15 additional SNPs (that either could not be designed for the OPA) from our 'top hits', *PODXL, ITGA6* and *MMP3*, by iPLEX.. This iPLEX is currently being designed and tested, and we expect genotyping to be complete within a month. Additional iPLEXes may be designed after Task 4 (validation) has been completed to more fully cover any genes in which we obtain independent validation of our results.

2. *statistical analysis of test set*

Preliminary analyses have been conducted from the OPA data, while the iPLEX data is pending. The main purpose of these preliminary analyses was to generate a list of SNPs for the Ovarian Cancer Association Consortium (OCAC) to genotype in the next three months for further validation. OCAC was founded in 2005 and now is comprised of 21 groups from Australia, Europe and America, with DNA and epidemiological data from ~4500 cases and 6500 controls (Gayther et al., 2007; Pearce et al., 2008; Ramus et al, in press). All analyses will be repeated when the iPLEX data are available.

All statistical analyses were conducted using the PLINK v0.99 Whole Genome Association Analysis toolset (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell et al., 2007). Of the 1536 SNPs genotyped using the using the Illumina Goldengate Assay, genotype data available for analysis consisted of a 1292 SNPs in a total of 1839 individuals following exclusions according to pre-determined quality control standards. Further quality control at the analytical level imposed by PLINK resulted in the exclusion of one SNP which failed the PLINK threshold of >10% of individuals with no genotype data, and three SNPs with a minor allele frequency (MAF) of <1%. Of the 1839 individuals with genotype data, three individuals were excluded by PLINK from all analyses because <10% of markers were successfully genotyped for these individuals. The final PLINK analysis data set consisted of a total of 1836 individuals for which genotype data on 1286 SNP were available. Summary statistics were obtained for each SNP on the frequency of missing genotype data among cases and controls as well as a comparison of missingness between cases and controls using the Fisher's exact test. A total of 37 (2.9%) SNPs had significantly different frequencies of missing genotype data between cases and controls (p<0.05).

Deviations from expected Hardy Weinberg (HW) proportions were analyzed using the Fisher's exact test and minor allele frequencies (MAFs) were also estimated for all SNPs. A basic allelic association test for ovarian cancer and each SNP was conducted comparing allele frequencies in cases and controls. The odds ratio (ORs) and 95% confidence intervals (CI) generated by this analysis represents the risk of ovarian cancer associated with the minor allele (m) for each SNP, and the unadjusted p-values were derived from 2 x 2 tables of ovarian cancer (cases vs. controls) by allele (m vs. M) using the chi-square test on 1 degree of freedom (df). Additional tests for allelic association for each SNP were implemented in PLINK included the Cochran-Armitage Trend test (1df), the general genotypic association test (2df) of ovarian cancer (cases vs. control) by genotype (mm vs. Mm vs. MM), the dominant gene association test (1df) of ovarian cancer (cases vs. controls) by dominant genotype (mm/Mm vs. MM), and the recessive gene association test (1df) of ovarian cancer (cases vs. controls) by recessive genotype (mm vs. Mm/MM).

Table 2 lists SNPs that had a P(trend) < 0.05 after applying the following exclusion criteria: SNPs with at least one failed duplicate, SNPs with a significantly different proportion of missing genotype data between cases and controls (P_{Miss} <0.05), SNPs not conforming to HW proportions (P_{HWE} <0.05) for either cases, controls or both, and SNPs with no significant trend in allelic dose response (P_{Trend} >0.05). From this list, we further estimated which SNPs are likely to be the best predictors of ovarian cancer (PPV) according to the p-values derived from the most robust test for allelic association i.e P_{Trend}, the power of the study to detect this association, and the prior probability of 0.0001. We will select SNPs for validation in Task 4 from this list.

Table 3 is a subset of the most highly ranked SNPs (by P (trend) value) from Table 2, with their Positive Predictive Values, that we proposed for validation to the whole of the Ovarian Cancer Association Consortium. The decision on 22-2-2008 was that OCAC would genotype four of these SNPs in *PODXL, ITGA6* and *MMP3* (2 SNPs) before the middle of June, after which we will get the data for analysis.

Table 2: SNPs with ovarian cancer risk estimates (P(trend) < 0.05)

Gene Symbol	CHR	SNP	Minor Allele	Major Allele	MAF$_{Controls}$	$^a P_{allelic}$	$^a OR_{allelic}$ (95% CI			P_{HWE}	P_{Trend}
ADAM8	10	rs1573041	A	G	0.2052	0.01372	1.23	(1.04	- 1.44)	0.5101	0.01465
CCL13	17	rs3136675	A	G	0.02461	0.0417	1.49	(1.01	- 2.20)	0.282	0.04464
CD44	11	rs1425802	G	A	0.2202	0.03048	1.19	(1.02	- 1.39)	0.6275	0.02788
CD44	11	rs10836342	C	G	0.3312	0.03238	0.85	(0.74	- 0.99)	0.4489	0.03237
CD44	11	rs2295756	G	A	0.3802	0.0411	0.86	(0.75	- 0.99)	0.2425	0.04051
CSF1	1	rs1999713	G	A	0.3423	0.003587	1.23	(1.07	- 1.41)	0.1905	0.003919
CTSK	1	rs4379678	G	A	0.07229	0.03942	1.29	(1.01	- 1.65)	0.6949	0.03758
DDR2	1	rs6693632	G	A	0.02984	0.01119	1.57	(1.11	- 2.22)	0.7495	0.01097
DDR2	1	rs6702820	G	A	0.2318	0.04228	0.84	(0.72	- 0.99)	0.3161	0.04209
EIF4EBP2	10	rs10999326	C	G	0.2695	0.0468	0.85	(0.73	- 1.00)	0.6414	0.04977
FGF2	4	rs17473132	A	G	0.06348	0.008271	1.41	(1.09	- 1.81)	1	0.007884
FGF2	4	rs167428	G	A	0.2524	0.02023	1.20	(1.03	- 1.39)	1	0.02027
FLT3LG	19	rs3826717	G	A	0.08398	0.009292	1.35	(1.08	- 1.69)	0.4041	0.009067
FN1	2	rs1250229	A	G	0.2768	0.01475	0.83	(0.71	- 0.96)	0.7127	0.01259
H1F0	22	rs763137	A	G	0.1163	0.01626	1.28	(1.05	- 1.55)	0.9027	0.0149
IFI16	1	rs1057024	G	A	0.1321	0.01687	1.26	(1.04	- 1.52)	0.2356	0.01753
IGFBP4	17	rs2245333	G	A	0.3212	0.01865	0.84	(0.72	- 0.97)	0.3316	0.0186
IGFBP5	2	rs11575194	A	G	0.03886	0.03972	1.39	(1.01	- 1.91)	0.5256	0.04155
IL1R1	2	rs3917332	T	A	0.2122	0.02175	0.82	(0.69	- 0.97)	0.2363	0.02216
ITGA6	2	rs13027811	G	A	0.1201	0.000828	0.68	(0.54	- 0.85)	0.8684	0.000857
ITGAV	2	rs11902171	G	C	0.2663	0.02169	1.19	(1.03	- 1.38)	0.2815	0.02057
ITGAV	2	rs3768787	G	A	0.216	0.03293	1.19	(1.01	- 1.39)	0.1721	0.03342
MMP1	11	rs7945189	A	G	0.09291	0.0284	1.28	(1.03	- 1.59)	0.9621	0.0319
MMP1	11	rs514921	G	A	0.3049	0.03536	0.85	(0.73	- 0.99)	0.8127	0.03431
MMP14	14	rs12050397	T	A	0.1714	0.03383	0.82	(0.68	- 0.98)	0.2553	0.03425
MMP3	11	rs522616	G	A	0.2319	0.001178	0.76	(0.64	- 0.90)	0.9314	0.001184
MMP3	11	rs650108	A	G	0.2763	0.01045	0.82	(0.70	- 0.95)	0.7159	0.01078
MMP7	11	rs17098236	A	G	0.09111	0.01864	0.74	(0.57	- 0.95)	0.2455	0.01673
MMP7	11	rs7935378	G	A	0.1668	0.03997	0.82	(0.68	- 0.99)	0.0709	0.03543
OSMR	5	rs10040172	G	A	0.185	0.008309	0.78	(0.65	- 0.94)	0.5488	0.009422
OSMR	5	rs2278324	A	C	0.1979	0.01515	0.80	(0.67	- 0.96)	0.9428	0.01647
OSMR	5	rs357287	C	A	0.3076	0.03182	0.85	(0.73	- 0.99)	0.9246	0.03172

Gene Symbol	CHR	SNP	Minor Allele	Major Allele	MAF$_{Controls}$	$^aP_{allelic}$	$^aOR_{allelic}$ (95% CI			P_{HWE}	P_{Trend}
PANX1	11	rs1540177	A	G	0.4247	0.02926	0.86	(0.75	- 0.98)	0.564	0.02973
PLOD2	3	rs1707469	C	A	0.3397	0.00559	1.22	(1.06	- 1.40)	0.3156	0.006242
PLOD2	3	rs1512900	C	G	0.4948	0.0132	0.84	(0.74	- 0.96)	1	0.01294
PODXL	7	rs1013368	G	A	0.338	0.000113	1.32	(1.14	- 1.51)	1	0.000104
PODXL	7	rs3735035	A	G	0.4983	0.03514	0.87	(0.76	- 0.99)	0.4138	0.03441
PODXL	7	rs1477250	G	A	0.5056	0.03938	0.87	(0.76	- 0.99)	0.9603	0.0385
PODXL	7	rs11768640	A	G	0.2211	0.04592	1.17	(1.00	- 1.38)	0.2797	0.04391
PTEN	10	rs34370136	A	G	0.05174	0.0307	1.36	(1.03	- 1.80)	0.9606	0.0314
PTTG1	5	rs7700446	A	G	0.1721	0.0132	0.79	(0.65	- 0.95)	0.923	0.01533
SAT	23	rs873637	A	G	0.06727	0.0393	1.30	(1.01	- 1.67)	0.2373	0.03851
SOX9	17	rs6501522	A	G	0.02119	0.006294	1.74	(1.16	- 2.60)	0.7298	0.006677
SPARC	5	rs3756631	T	A	0.125	0.01084	1.28	(1.06	- 1.56)	0.6536	0.01146
TERT	5	rs7726159	A	C	0.3159	0.006433	1.22	(1.06	- 1.40)	0.6199	0.00675
TERT	5	rs11133719	A	G	0.1735	0.02645	0.81	(0.67	- 0.98)	0.8838	0.02479
TGFB2	1	rs10495098	A	C	0.3786	0.01337	1.19	(1.04	- 1.36)	0.5754	0.01395
THBS4	5	rs17879514	A	G	0.06649	0.03852	0.73	(0.55	- 0.98)	0.5443	0.03791
TIMP3	22	rs5754289	A	G	0.1745	0.007263	1.26	(1.06	- 1.49)	0.5418	0.007529
TIMP3	22	rs130290	A	G	0.09213	0.0184	0.74	(0.57	- 0.95)	0.747	0.01845
VDR	12	rs11574139	A	T	0.04066	0.0284	0.65	(0.44	- 0.96)	1	0.02676
VEGF	6	rs3025040	A	G	0.1332	0.04292	1.22	(1.01	- 1.47)	0.09188	0.04018
WNT5A	3	rs590386	A	G	0.0924	0.04168	1.26	(1.01	- 1.56)	0.4913	0.04394

a   Odds ratios, 95% CI and p-values are derived from the allelic test for association (m vs. M) using χ^2 test on 1 df

Table 3: SNPs proposed for validation by the OCAC

Gene symbol	SNP	Prior	[a]Alpha	[b]Power	[c]PPV
PODXL	rs1013368	0.0001	0.0001037	0.51	33.1%
ITGA6	rs13027811	0.0001	0.0008566	0.40	4.5%
MMP3	rs522616	0.0001	0.001184	0.55	4.4%
TERT	rs7726159	0.0001	0.00675	0.98	1.4%
TIMP3	rs5754289	0.0001	0.007529	0.90	1.2%
FGF2	rs308441	0.0001	0.006472	0.73	1.1%
FGF2	rs17473132	0.0001	0.007884	0.84	1.1%
SSA1	rs4144331	0.0001	0.004018	0.35	0.9%
SOX9	rs6501522	0.0001	0.006677	0.55	0.8%
PLOD2	rs1512900	0.0001	0.01294	1.00	0.8%
MMP3	rs650108	0.0001	0.01078	0.43	0.4%
CSF1	rs1999713	0.0001	0.003919	0.15	0.4%
FGF2	rs167428	0.0001	0.02027	0.73	0.4%
PLOD2	rs1707469	0.0001	0.006242	0.22	0.3%
TIMP3	rs130290	0.0001	0.01845	0.38	0.2%
PTTG1	rs17057781	0.0001	0.003465	0.06	0.2%
TERT	rs11133719	0.0001	0.02479	0.30	0.1%
PTTG1	rs7700446	0.0001	0.01533	0.17	0.1%

a: P-values from Cochran-Armitage test for allelic trend
b: Power of the study to detect the association
c: Positive predictive value

Task 4. **Genotyping of the replication set and statistical analysis of replication set (months 22-32)**

1. *genotyping 1200 cases and 3600 controls by Mass Array for 45-60 SNPs in 30-plexes, significantly associated with ovarian cancer risk in the test set (P< 0.001)*

We are currently collecting ovarian case-control DNAs from six members of OCAC - SEARCH (PI: Paul Pharoah), MALOVA (Estrid Hogdall), FROCS (Alice Whittemore), UKOPS (Simon Gayter), University of Southern California (Leigh Pearce) and Mayo Clinic (Ellen Goode) in order to genotype the most significant SNPs from our first phase. We anticipate receiving ~2000 case and ~4000 control DNAs from these studies for the replication set. This will be done by iPLEX, and so we plan to test the most significant 25-30 SNPs (depending on how many will fit into the multiplex). The threshold P value will be ~0.01 because only two SNPs (in *PODXL* and *ITGA6*) fall under our original, more stringent threshold of 0.0001 but there is no incremental cost in terms of DNA amounts, and very little financial cost, to genotyping 25-30 SNPs in a single iPLEX reaction, instead of only two as originally planned. We have requested by DNAs by early April, so anticipate that this will be completed by the end of May.

2. *statistical analysis*

This will be performed in June and July 2008.

Task 5. **DHPLC to identify putative functional SNPs in genes associated with**

serous invasive ovarian cancer in both the test and replication set (months 25-35)

1. *design of DHPLC primers*
2. *DHPLC of coding and conserved regulatory regions of ~5 genes in 94 moderate familial risk ovarian cancer cases*

This will not commence until Task 4 has been completed.

Task 6. **Functional evaluation of putative rSNPs (months 28-36)**

This will not commence until Task 5 has been completed.

Task 7. **Manuscript preparation (months 32-)**

KEY RESEARCH ACCOMPLISHMENTS

We have genotyped 2138 samples (773 invasive, serous cases plus 1365 controls) for 1536 tagging, non-synonymous and miRNA binding site SNPs in 174 genes. Following Quality Control exclusions, the final dataset comprised 1839 samples (675 cases and 1164 controls) with genotype information for 1292 SNPs in 174 genes. We are using P(trend) values to select 25-30 SNPs for independent validation in ~2000 cases and ~4000 controls from other sites. Four of these SNPs will be genotyped by the whole of OCAC (~4500 cases and 6500 controls).

REPORTABLE OUTCOMES

Abstract presented at the AACR meeting on 'Approaches to complex pathways in molecular epidemiology' in Albuquerque in May 2007.

CONCLUSION

Progress is satisfactory, but there are no validated conclusions to report yet. However, we are encouraged by our analyses to date, and in particular the significance of the finding for a podocalyxin-like (*PODXL*) SNP for which the P(trend) = 0.0001, with a Positive Predictive Value of 33 %, and a Homozygote OR = 1.75 (95% CI 1.28-3.38). If this replicates in our validation phase, it will have important implications for the etiology, and perhaps prognosis, of ovarian cancer. *PODXL* (podocalyxin-like protein) maps to the 7q32-q33 region that has shown strong linkage to aggressive prostate cancer (Neville et al., 2002) and encodes a mucin-like extracellular matrix protein involved in cell adhesion. The mechanism by which podocalyxin increases cancer aggressiveness remains poorly understood. Sizemore et al (2007) showed that overexpression of podocalyxin in MCF7 breast and PC3 prostate cancer cell lines increased their *in vitro* invasive and migratory potential and led to increased expression of matrix metalloproteases 1 (MMP1) and 9 (MMP9), suggesting that podocalyxin may be involved in the metastatic phenotype and poor outcome. Somasiri et al (2004) found that podocalyxin is highly overexpressed in a subset of invasive breast carcinoma, and that podocalyxin was an independent predictor of poor outcome. Schopperle et al (2003) recently identified GP200 as a testicular tumour form of podocalyxin. PODXL has also been identified as a

candidate gene involved in primordial follicle formation in gene expression profiling studies of mouse ovary development.

REFERENCES

Gayther et al on behalf of the Ovarian Cancer Association Consortium. Tagging Single Nucleotide Polymorphisms in Cell Cycle Control Genes and Susceptibility to Invasive Epithelial Ovarian Cancer. *Cancer Research* 67: 3027-3035 (2007)

Neville PJ, Conti DV, Paris PL, et al: Prostate cancer aggressiveness locus on chromosome 7q32-q33 identified by linkage and allelic imbalance studies. Neoplasia 4:424-31, 2002

Pearce et al on behalf of the Ovarian Cancer Association Consortium (OCAC). Progesterone Receptor Variation and Risk of Invasive Epithelial Ovarian Cancer: Results from the Ovarian Cancer Association Consortium Pooled Analysis. *British Journal of Cancer* 98:282-288 (2008)

Purcell S, Neale B, Todd-Brown K, et al: PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-75, 2007

Ramus et al on behalf of the Ovarian Cancer Association Consortium (OCAC) Consortium Analysis of Seven Candidate SNPs for Ovarian Cancer. *International Journal of Cancer* (in press0

Sizemore S, Cicek M, Sizemore N, et al: Podocalyxin increases the aggressive phenotype of breast and prostate cancer cells in vitro through its interaction with ezrin. Cancer Res 67:6183-91, 2007

Somasiri A, Nielsen JS, Makretsov N, et al: Overexpression of the anti-adhesin podocalyxin is an independent predictor of breast cancer progression. Cancer Res 64:5068-73, 2004

Schopperle WM, Kershaw DB, DeWolf WC: Human embryonal carcinoma tumor antigen, Gp200/GCTM-2, is podocalyxin. Biochem Biophys Res Commun 300:285-90, 2003

APPENDICES

None